



UNCLASSIFIED

Ref:
Issue/Revision:
Date:RX-RP-53-5690
1/0
OCT. 30, 2013

Automatic Target Cueing (ATC)

Task 1 Report – Literature Survey on ATC

October 30, 2013

Prepared by:
Stephen Se

MDA Systems Ltd.
13800 Commerce Parkway
Richmond, BC, Canada
V6V 2J3

PWGSC Contract Title:	Automatic target cueing and facial recognition in visible & infrared spectrum for military operations
MDA Project Title:	6484 - ATC
MDA Document Number:	RX-RP-53-5690
Contract No.:	W7701-135590/001/QCL
Project Duration:	August 2 2013 – March 31 2014
Contract Scientific Authority:	Dr. Philips Laou (418) 844-4000 x4039

DRDC-RDDC-2014-C173

The scientific or technical validity of this Contract Report is entirely the responsibility of the contractor and the contents do not necessarily have the approval or endorsement of Defence R&D Canada.

© Her Majesty the Queen in Right of Canada, as represented by the Minister of National Defence, 2014

Use, duplication, or disclosure of this document or any of the information contained herein is subject to the restrictions on the title page of this document.



UNCLASSIFIED

Ref:
Issue/Revision:
Date:

RX-RP-53-5690
1/0
OCT. 30, 2013

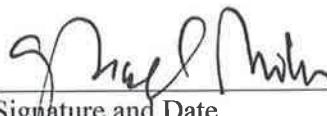
Prepared By:

Stephen Se

Stephen Se Oct 30, 2013
Signature and Date

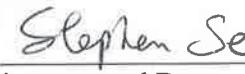
Checked By:

Mohsen Ghazel

Mohsen Ghazel Oct 30, 2013
Signature and Date

Project Manager:

Stephen Se

Stephen Se Oct 30, 2013
Signature and Date

MacDonald, Dettwiler and Associates Ltd
13800 Commerce Parkway
Richmond, BC, Canada
V6V 2J3

Use, duplication, or disclosure of this document or any of the information contained herein is subject to the restrictions on the title page of this document.



UNCLASSIFIED

Ref:
Issue/Revision:
Date:RX-RP-53-5690
1/0
OCT. 30, 2013

CHANGE RECORD

ISSUE	DATE	PAGE(S)	DESCRIPTION
1/0	Oct. 30, 2013	All	First Issue

(v)

Use, duplication, or disclosure of this document or any of the information contained herein is subject to the restrictions on the title page of this document.

Ref: RX-RP-53-5690
Issue/Revision: 1/0
Date: OCT. 30, 2013

UNCLASSIFIED



THIS PAGE INTENTIONALLY LEFT BLANK

(vi)

Use, duplication, or disclosure of this document or any of the information contained herein is subject to the restrictions on the title page of this document.

TABLE OF CONTENTS

1	INTRODUCTION	1-1
1.1	Background.....	1-1
1.2	Project Objectives.....	1-1
1.3	Task 1 Objectives	1-2
1.4	Scope.....	1-2
2	LITERATURE SURVEY.....	2-1
2.1	ATC Review Papers	2-1
2.2	Generic ATC Pipeline	2-5
2.2.1	Pre-processing.....	2-5
2.2.2	Detection.....	2-7
2.2.3	Feature Extraction.....	2-8
2.2.4	Classification/Recognition.....	2-14
2.2.5	Object Tracking	2-16
2.3	End-to-End ATC Systems.....	2-20
2.3.1	Infrared ATC Systems	2-21
2.3.2	Visible ATC Systems	2-21
2.3.3	Multi-Sensor ATC Systems	2-23
2.4	ATC Performance Evaluation.....	2-23
2.5	COTS ATC Software/SDK.....	2-26
2.5.1	COTS Software.....	2-26
2.5.2	SDK Applicable to ATC Development	2-27
2.6	Concluding Remarks.....	2-27
3	REFERENCES.....	3-1

Ref: RX-RP-53-5690
Issue/Revision: 1/0
Date: OCT. 30, 2013

UNCLASSIFIED



THIS PAGE INTENTIONALLY LEFT BLANK

(viii)

Use, duplication, or disclosure of this document or any of the information contained herein is subject to the restrictions on the title page of this document.

LIST OF FIGURES

Figure 2-1	Generic ATC Pipeline	2-5
Figure 2-2	Skeleton Examples (a) Distance Transform (b) Iterative Thinning	2-10
Figure 2-3	Example of Wide Baseline Matching Between Two Images With SIFT	2-11
Figure 2-4	FAST Feature Detection Example (Source: [P-11]).....	2-11
Figure 2-5	Local Self-Similarity Descriptor Example at an Image Pixel (Source: [P-13])	2-12
Figure 2-6	Detection of a Person Raising Both Arms (a) A Hand-sketched Template (b) Detected Locations in Other Images (Source: [P-13])	2-13
Figure 2-7	HOG Cue Mainly on Silhouette Contours (Head, Shoulders & Feet) (Source: [P-10])	2-15
Figure 2-8	Test Image at Four Scales (a) Original Size (b) Factor-2 Reduction (c) Factor-4 Reduction (d) Factor-8 Reduction (Source: [P-19])	2-15
Figure 2-9	Taxonomy of Tracking Methods (Source: [P-21])	2-18
Figure 2-10	Thermal Images of Human Activities (Source: [P-37]).....	2-25
Figure 2-11	Thermal Images of Two-Hand Object Identification (Source: [P-37])	2-25

Ref: RX-RP-53-5690
Issue/Revision: 1/0
Date: OCT. 30, 2013

UNCLASSIFIED



THIS PAGE INTENTIONALLY LEFT BLANK

(x)

Use, duplication, or disclosure of this document or any of the information contained herein is subject to the restrictions on the title page of this document.

LIST OF TABLES

Table 2-1	Selected Papers on ATC Review	2-1
Table 2-2	Selected Papers on Pre-processing	2-6
Table 2-3	Comparison of Selected Papers on Pre-processing	2-6
Table 2-4	Selected Papers on Detection	2-7
Table 2-5	Comparison of Selected Papers on Detection.....	2-8
Table 2-6	Selected Papers on Feature Extraction	2-8
Table 2-7	Comparison of Selected Papers on Feature Extraction	2-13
Table 2-8	Selected Papers on Classification/Recognition	2-14
Table 2-9	Comparison of Selected Papers on Classification/Recognition	2-16
Table 2-10	Selected Papers on Object Tracking.....	2-17
Table 2-11	Comparison of Selected Papers on Object Tracking.....	2-19
Table 2-12	Selected Papers on ATC Systems for Infrared Spectrum.....	2-20
Table 2-13	Selected Papers on ATC Systems for Visible Spectrum	2-20
Table 2-14	Selected Papers on Multi-Sensor ATC Systems.....	2-20
Table 2-15	Comparison of ATC Systems for Infrared Spectrum	2-21
Table 2-16	Comparison of ATC Systems for Visible Spectrum.....	2-22
Table 2-17	Comparison of Multi-Sensor ATC Systems	2-23
Table 2-18	Selected Papers on ATC System Performance Evaluation.....	2-23
Table 2-19	Comparison of COTS Software Applicable to ATC	2-28
Table 2-20	Comparison of SDK Applicable to ATC Development	2-31

Ref: RX-RP-53-5690
Issue/Revision: 1/0
Date: OCT. 30, 2013

UNCLASSIFIED



THIS PAGE INTENTIONALLY LEFT BLANK

(xii)

Use, duplication, or disclosure of this document or any of the information contained herein is subject to the restrictions on the title page of this document.

ACRONYMS AND ABBREVIATIONS

2D	Two Dimensional
3D	Three Dimensional
ALERT	Advanced Linked Extended Reconnaissance Targeting
ATC	Automatic Target Cueing
ATE	Assisted Target Engagement
ATR	Automatic Target Recognition
BRIEF	Binary Robust Independent Elementary Features
CAVIAR	Context Aware Vision using Image-based Active Recognition
CCTV	Closed Circuit Television
CLEAR	Classification of Events, Activities and Relationships
COTS	Commercial Off-The-Shelf
CPU	Central Processing Unit
CVC	Computer Vision Center
CVPR	Computer Vision and Pattern Recognition
DARPA	Defense Advanced Research Projects Agency
DLL	Dynamic Link Library
DOD	Department Of Defense
DRDC	Defence Research and Development Canada
ECCV	European Conference on Computer Vision
EO	Electro-Optical
ETH	Elgenoössische Technische Hachschule (Swiss Federal Institute of Technology)
FAST	Features from Accelerated Segment Test
FR	Facial Recognition
FSAR	Future Small Arms Research
GB	Giga Bytes
GHz	Giga Hertz
GLOH	Gradient Location and Orientation Histogram

(xiii)

GPU	Graphics Processing Unit
GSD	Ground Sample Distance
HOF	Histograms of Flows
HOG	Histograms of Oriented Gradients
Hz	Hertz
ICCV	International Conference on Computer Vision
ICVS	International Conference on Computer Vision Systems
IEEE	Institute of Electrical and Electronics Engineers
IJCV	International Journal of Computer Vision
INRIA	Institut National de Recherche en Informatique et en Automatique
IPP	Integrated Performance Primitives
IR	Infrared
IROS	Intelligent Robots and Systems
KLT	Kanada-Lucas-Tomasi
LADAR	Laser Detection and Ranging
LBP	Local Binary Pattern
LSS	Local Self-Similarity
LWIR	Long Wave Infra-Red
MDA	MDA Systems Ltd.
MEX	Matlab Executable
MIT	Massachusetts Institute of Technology
MPEG	Moving Picture Experts Group
NA	Not Available
NVESD	Night Vision and Electronic Sensors Directorate
OpenCV	Open Source Computer Vision Library
ORB	Oriented FAST and Rotated BRIEF
PCA	Principal Component Analysis
PETS	Performance Evaluation of Tracking and Surveillance
PLSS	A low-dimensional variant of LSS using PCA



UNCLASSIFIED

Ref:
Issue/Revision:
Date:RX-RP-53-5690
1/0
OCT. 30, 2013

RANSAC	Random Sample Consensus
ROC	Receiver Operating Characteristic
RSJ	The Robotics Society of Japan
SDK	Software Development Kit
SIFT	Scale Invariant Feature Transform
SNR	Signal-to-Noise Ratio
SPIE	Society of Photo-Optical Instrumentation Engineers
START	Scoring, Truthing, And Registration Toolkit
SURF	Speeded Up Robust Features
SVM	Support Vector Machine
TDP	Technical Demonstration Project
TPP	Targeting Task Performance
TUD	Technische Universitat Darmstadt
UAV	Unmanned Aerial Vehicles
US	United States
USC	University of Southern California
VIRAT	Video and Image Retrieval and Analysis Tool
VMTI	Video Moving Target Indication
XML	Extensible Markup Language

Ref: RX-RP-53-5690
Issue/Revision: 1/0
Date: OCT. 30, 2013

UNCLASSIFIED



THIS PAGE INTENTIONALLY LEFT BLANK

(xvi)

Use, duplication, or disclosure of this document or any of the information contained herein is subject to the restrictions on the title page of this document.

1 INTRODUCTION

1.1 Background

Under the mandate of Future Small Arms Research (FSAR) program, Defence Research and Development Canada (DRDC) will examine existing and future technologies for small arms capabilities with the objective of identifying technologies which could increase shot placement accuracy and reduce engagement time.

Automatic Target Cueing (ATC) is considered as one of the key enablers in future small arms technology. The goal is to assess feasibility and capability of Assisted Target Engagement (ATE) in small arms by combining ATC with electronic ignited ammunition and small arms weapons. It is believed that ATE may help improving shot placement and shortening engagement time in some situations.

1.2 Project Objectives

The objectives of this project are to conduct a feasibility study and provide software development support on Automatic Target Cueing (ATC) and Facial Recognition (FR) in visible and infrared spectrum for military operations.

The project consists of seven tasks: the first two tasks are related to literature survey of ATC and FR, the next two tasks are related to the study and evaluation of existing ATC and FR products, while the last three tasks are related to enhancing existing DRDC system and developing new ATC capabilities.

1.3 Task 1 Objectives

The objective of Task 1 is to conduct a review of existing literature on ATC methodologies and technologies based on imagery, in order to determine the feasibility of performing highly accurate ATC from visible and infrared spectrum in military operations. The military operations include (i) ATC in small arms operations at standoff distance up to 600m; and (ii) short range ATC at standoff distance below 100m.

This task consists of two sub-tasks:

- Task 1.1: Conduct a literature survey, identify, describe and interpret the key characteristics of ATC methodologies and technologies in infrared spectrum
- Task 1.2: Conduct a literature survey, identify, describe and interpret the key characteristics of ATC methodologies and technologies in visible spectrum

1.4 Scope

This report fulfils Task 1 milestone of this contract and contains the following elements:

- Literature survey of ATC methodologies and technologies
- Literature survey of ATC systems and performance evaluation
- Review of COTS ATC software/SDK

2

LITERATURE SURVEY

This Section presents a literature survey of the various ATC technologies and methodologies in both infrared and visible spectrum. We start by reviewing several survey papers on ATC. We then describe a generic ATC pipeline and its key processing steps, followed by reviewing papers relevant to each of those steps. Next, we review various papers on ATC end-to-end systems as well as ones related to ATC performance evaluation. Finally, we review a number of COTS software and SDK applicable to ATC. The focus here is on small arms applications, in particular human target detection at long range operation.

2.1 ATC Review Papers

ATC has become increasingly important in modern defense strategy because it permits precision strikes against certain tactical targets with reduced risk and increased efficiency, while minimizing collateral damage. By making computer detect and recognize targets automatically, the workload of the soldiers can be reduced and the accuracy and efficiency of the weapons can be improved. Table 2-1 lists the selected ATC review papers, which are specifically for military applications.

Table 2-1 Selected Papers on ATC Review

#	Paper Title	Authors	Source	Year
P-1	Aided and Automatic Target Recognition Based Upon Sensory Inputs From Image Forming Systems	J.A. Ratches, C.P. Walters, R.G. Buser, B.D. Guenther	IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(9)	1997
P-2	Review of Current Aided/Automatic Target Acquisition Technology for Military Target Acquisition Tasks	J.A. Ratches	Optical Engineering, 50(7)	2011

[P-1] reviews 10 years of Automatic Target Recognition (ATR) research conducted at United States (US) Army Laboratories. The key points from the paper are as follows:

- Metrics for ATR algorithms evaluation:
 - Signal-to-Noise Ratio (SNR): A ratio of signal power to the noise power
 - Receiver Operating Characteristic (ROC) curves: A plot to show the probability of detection against the probability of false alarm at different thresholds
 - Confusion matrix: 2D array to indicate the identity assigned by the ATR system versus the ground truth
 - Consistency: A measure of how often an ATR algorithm gives the same result for successive image frames of the same scene, in the presence of noise
- Algorithm development progress
 - Algorithms in the early 1980s were heuristic, as the target detection was based on some sort of threshold, determined by the contrast of an object compared to the local background. Detection in low clutter did not exceed 70%, and recognition was little better than random guessing. False alarm rates were mostly unacceptable.
 - In the late 1980s and into the 1990s, a new generation of algorithms was developed, using knowledge-based systems or template-matching approaches. Each match between a region of interest and a template results in a score that can be subjected to a thresholding procedure for false alarm reduction. Performance evaluation of these systems has shown a significant improvement. Detection in low to medium clutter increased to 80% but the false alarm rate is still high.
- Hardware platform trend
 - Early history of military ATR emphasized hardware development to supply great computing power. For example, Aladdin was a Defense Advanced Research Projects Agency (DARPA) project to develop parallel processor in a miniature, modular, high-density package, which can perform ATR functions on a 128×128 pixel image at 30 frames per second.
 - However, present and future trend is through leverage of commercial hardware evolution and concentration on more processing capability with massively parallel architectures.
- Measured performance trend
 - Over 75 ATR implementations have been evaluated by the US Army since early 1980s. Documented and quantitative data has shown improvement over the years, but much of the data appeared in the classified literature only.
 - Probability of detection rises with the SNR and approaches a limiting value, and it was noticed that the knee is around an SNR of 5, independent of algorithms. Clutter has a severe effect on false alarm rate.

- While an ATR algorithm may exhibit performance levels below the human in terms of probability of detection, it is tireless and it is many times faster than a human.
- Latest techniques under consideration by the military research community
 - Multi-sensor fusion
 - Promising approach to increase performance is through the use of independent information, i.e. multi-sensors or integration of spatial and temporal information.
 - Multi-sensor could include other sensor modalities or non-imaging sources. For example, range data from a laser range-finder or a radar can provide size of the target, so that objects, which are too small or too large can be rejected as false alarms. Recent tests by the US Army have shown that sensor fusion (infrared and radar) can provide an order of magnitude reduction in false alarm rates over single-sensor performance.
 - Model-based algorithms
 - Model-based algorithms contain libraries of models of the targets for scenarios of interest. Images from the sensor are compared to the library models which are coupled with environmental effects. A canonical database of targets is essential to build the model templates.

[P-2] is a more recent literature review describing the work at the US Army Laboratories on ATR. It discusses the key challenges and potential technical approaches that could enable new advancements in military-relevant performance. The key points from the paper are as follows:

- Military importance
 - ATR is an extremely important technology for military operations but has not yet realized its full tactical promise. For weapon systems, the primary value is the reduction in engagement time for target acquisition. The rapid acquisition and servicing of targets increase lethality and survivability of the weapon platform and soldier.
 - US Army, Navy and Air Force are all pursuing ATR for intelligence, surveillance, reconnaissance, target acquisition, wide-area search & track, etc. Ground-to-ground missions are extremely difficult due to high chance of encountering clutter, compared to air-to-ground missions.
- Algorithm development trend
 - Various algorithms have been proposed including statistical, shape-based (template/model), moving target indicator, increased dimensionality (e.g. 3D LADAR), hyper-spectral, multi-spectral, etc.
 - Multi-sensor approaches have been tried, including:
 - Multi-sensor where more than one sensor is looking at the same target
 - Multi-look where one sensor gets several looks at the target from different views
 - Multi-mode fusion where sensor of different modalities sense the target (e.g. acoustic and Electro-Optical (EO))

- Performance evaluation
 - Three bottom-line figures of merit for ATR evaluation are ROC curves, confusion matrices and time.
 - The lack of unclassified dataset has been addressed recently, with the release of a specially gathered unclassified imagery. Over 300 GB of imagery data (infrared and visible) of tactical vehicles, civilian vehicles and people in realistic tactical scenes with ground-truth is now available.
 - Testing with simulated imagery has shown that while the detection probabilities are quite comparable between synthetic and realistic imagery, the false alarm rate was much different, as the synthetic noise generation can be significantly different from the true sensor noise characteristics.
- State-of-the-art ATR still produces unacceptable number of false alarms in highly cluttered background, due to the following key challenges:
 - False alarms: The primary operational limitation for ATR system is false alarm rate, which can be so high that the operator will turn off the system. False alarm not only causes the operator to spend excessive time checking them, firing at a false target will give away the position of the firing platform and make it a target of counter-fire.
 - Clutter: A primary limitation of ATR is the lack of an understanding of clutter and a reliable clutter model that can quantify the scene difficulty. The ultimate clutter metric must contain some target conspicuity factor. A clutter metric that is simply a function of signal-to-noise ratio will not reflect the true dependency of performance on real-world clutter.
 - Target variability: Target appearance can vary a lot under different environmental, operational and background conditions. Moreover, camouflage, concealment and decoys increase the target dimensional space significantly.
- Promising approaches
 - Shape-based approaches to ground-to-ground scenarios have shown to give useful performance in low to medium clutter. Approaches with some success include change detection and moving target indication.
 - ATR from airborne sensor platform has shown better performance than ground scenarios, as recognizing an overhead view is not as complex and not as easily confused with clutter.
 - Promising sensor approaches include 3D LADAR and multi-spectral/hyper-spectral sensors. However, such systems require increased system complexity and cost.
 - Aided target recognition will mature more rapidly than ATR, by offloading the higher level decisions to human. Aided target recognition will provide an order of magnitude improvement in target acquisition times than human alone.

2.2 Generic ATC Pipeline

Figure 2-1 shows a generic ATC pipeline adapted from [P-2], consisting of the following key processing steps:

1. Pre-processing step enhances and de-noises the imagery.
2. Detection step indicates whether any foreground object is present.
3. Feature Extraction step extracts features from the image regions.
4. Classification/Recognition step discerns a type of object, e.g. a person versus a car.
5. Object Tracking step tracks the recognized object over frames.
6. Identification step discerns a specific person in the FSAR context. Face recognition is described in detail in the Task 2 report, and hence it will not be discussed further here.

In the following sub-sections, various relevant papers of these key processing steps are reviewed and compared based on the FSAR criteria.

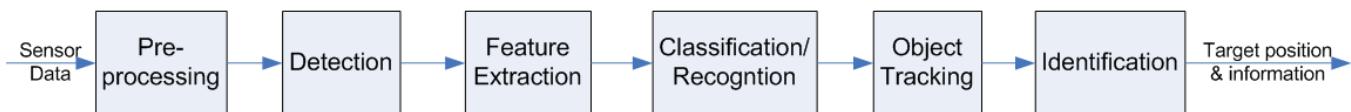


Figure 2-1 Generic ATC Pipeline

2.2.1 Pre-processing

The purpose of pre-processing is to enhance the input imagery data, so that the subsequent processing steps can generate better results. In general, any pre-processing algorithms that can improve the stabilization/resolution/contrast of the input data would be applicable to FSAR. While there are many papers in the literature on image pre-processing, the papers reviewed here, as listed in Table 2-2, are specifically for target acquisition and ATC-related applications.

- [P-3] investigated the effects of using several medical imaging enhancement techniques such as contrast/edge enhancement to increase the detectability of targets in the urban terrain.
- [P-4] restores long-distance thermal videos using a blind image de-convolution method to correct for the atmospheric degradations that may reduce the quality of such videos and hence the ability to acquire moving targets automatically.
- [P-5] evaluates the target acquisition performance improvement by super-resolution processing, which helps with increasing the sampling, removal of aliasing and reduction of fixed-pattern noise.

The key findings of these papers are listed in Table 2-3 for comparison, with respect to the various FSAR criteria. Although only a limited number of pre-processing techniques are reviewed, this shows that most pre-processing techniques that can improve the input image quality would also boost the ATC performance.

Table 2-2 Selected Papers on Pre-processing

#	Paper Title	Authors	Source	Year
P-3	Effect of Image Enhancement on the Search and Detection Task in the Urban Terrain	N. Devitt, S. Moyer, S. Young	SPIE Vol. 6207	2006
P-4	Improvement of Automatic Acquisition of Moving Objects in Long-Distance Imaging by Blind Image Restoration	O. Haik, Y. Yitzhaky	SPIE Vol. 6737	2007
P-5	IR System Field Performance with Superresolution	J. Fanning, J. Miller, J. Park, G. Tener, J. Reynolds, P. O'Shea, C. Halford, R. Driggers	SPIE Vol. 6543	2007

Table 2-3 Comparison of Selected Papers on Pre-processing

Paper	Pre-processing Techniques	Input Data	Low Resolution/ Long Range?	Sensor	Experimental Setup	Results
[P-3]	<ul style="list-style-type: none"> • Contrast enhancement • Edge enhancement • Multi-scale edge domain 	Still images	Yes	IR	Human observers visually compare original and enhanced images in terms of probability of detection & time to detect target, for both day-time and night-time imagery.	<ul style="list-style-type: none"> • Only contrast enhanced IR night-time imagery show measureable improvement • Results for other techniques are inconclusive or insignificant
[P-4]	<ul style="list-style-type: none"> • Blind image restoration to correct for long range atmospheric degradation 	Video from fixed camera	Yes (examples up to 3km)	IR	Original & restored images are processed by computer algorithms (motion detection & tracking) and the results are compared.	<ul style="list-style-type: none"> • Substantial details are uncovered from restored imagery • Image restoration improves target acquisition for both human visual system & computerized applications
[P-5]	<ul style="list-style-type: none"> • Super-resolution • Super-resolution with de-blurring 	Video from moving camera	Yes (examples up to 1.2km)	IR	Human observers visually compare video with and without super-resolution in terms of probability of identification.	<ul style="list-style-type: none"> • Super-resolution produces significant performance increase • Improvement is more substantial for under-sampled sensors • Wiener filter de-blurring further improves the results slightly

2.2.2 Detection

Most video surveillance systems include a detection step to separate the moving objects from the background clutter, so that further analysis can be focused only on the foreground objects. This is also referred to as Video Moving Target Indication (VMTI). A key benefit of this step is to reduce the amount of data for the subsequent steps to improve throughput. While these foreground segmentation techniques would not work if the targets remain stationary in the scene, it is highly likely they would move around in typical battlefield scenarios, and would be detected by these techniques. Table 2-4 lists the selected papers on detection.

Table 2-4 Selected Papers on Detection

#	Paper Title	Authors	Source	Year
P-6	Video Moving Target Indication in the Analysts' Detection Support System	R. Jones, D.M. Booth, N.J. Redding	Australian Government Department of Defence DSTO-RR-0306	2006
P-7	Independent Component Analysis-based Background Subtraction for Indoor Surveillance	D.M. Tsai, S.C. Lai	IEEE Transactions on Image Processing, 18(1)	2009

[P-7] detects moving objects in an image sequence for indoor video surveillance, where the camera is stationary. Background subtraction is a popular technique for foreground segmentation. However, the background model needs to be updated continuously to compensate for illumination changes over time. A fast background subtraction scheme using independent component analysis is proposed, which is computationally as fast as simple image difference method and yet is highly tolerable to changes in room lighting.

Techniques suitable for VMTI for stationary camera, such as background subtraction, would not work for moving cameras in the FSAR application. For stationary cameras, apart from illumination changes or object movement caused by wind, pixel changes are due to moving objects. However, in the case of moving cameras, everything in the scene appears to be moving relative to the camera. The motion of the actual targets must be distinguished from the global motion in the scene, i.e. the goal is to find pixel movement not caused by the camera motion.

A feature-based approach has been developed in [P-6], with Kanada-Lucas-Tomasi (KLT) feature tracking and Random Sample Consensus (RANSAC) outlier removal, combined with background modeling and frame differencing. The solution provides size and positional information on a frame-by-frame basis for any moving targets in a video sequence from a moving camera. VMTI needs to run at near real-time to be of any operational value in the field. Suitable parallel non-specialized hardware is proposed to achieve near real-time performance.

Table 2-5 Comparison of Selected Papers on Detection

Paper	Input Data	Low Resolution/ Long Range?	Sensor	Real-time Processing?
[P-6]	Video from moving camera	Yes	Only visible results shown	Not currently. Expected to be real-time given suitable parallel non-specialized hardware
[P-7]	Video from fixed camera	Yes	Only visible results shown	Yes

2.2.3 Feature Extraction

Once the foreground objects have been detected, the next step is to extract features from those regions so that the features can be used to check whether any object of interest is present, i.e. human targets in the FSAR application.

Feature extraction has been an active research topic in the computer vision/image processing community for decades. There are numerous feature extraction techniques, ranging from basic ones such as edge and corner detector to more sophisticated ones. Table 2-6 lists the papers reviewed here, which include a few well-known references as well as references that have been identified by DRDC as promising. Most of these references describe generic feature extraction techniques, rather than specifically designed for human detection. [P-16] is a recent Ph.D. thesis on human detection, with a chapter on the state-of-the-art review, which will be summarized first. This is then followed by the review of the other papers.

Table 2-6 Selected Papers on Feature Extraction

#	Paper Title	Authors	Source	Year
P-8	Skeletons in Digital Image Processing	G. Klette	Computer Science Dept, University of Auckland CITR-TR-112	2002
P-9	Distinctive Image Features from Scale-Invariant Keypoints	D. Lowe	International Journal of Computer Vision (IJCV), 60(2)	2004
P-10	Histograms of Oriented Gradients for Human Detection	N. Dalal and B. Triggs	IEEE Conference on Computer Vision and Pattern Recognition (CVPR)	2005
P-11	Fusing Points and Lines for High Performance Tracking	E. Rosten, T. Drummond	International Conference on Computer Vision (ICCV)	2005
P-12	Machine Learning for High-speed Corner Detection	E. Rosten, T. Drummond	European Conference on Computer Vision (ECCV)	2006
P-13	Matching Local Self-Similarities across Images and Videos	E. Shechtman, M. Irani	IEEE Conference on Computer Vision and Pattern Recognition (CVPR)	2007
P-14	ORB: an Efficient Alternative to SIFT or SURF	E. Rublee, V. Rabaud, K. Konolige, G. Bradski	International Conference on Computer Vision (ICCV)	2011
P-15	Low-Dimensional Local Self-Similarity Descriptor for Image Matching	J. Liu, G. Zeng	Advances in Automation and Robotics Vol. 2, Lecture Notes in Electrical Engineering Vol. 123	2012
P-16	Human Detection from Images and Videos (Chapter 2)	D.T. Nguyen	Ph.D. Thesis, University of Wollongong	2012

[P-16] includes a state-of-the-art review of features and object representation for human detection. Features are categorized based on the aspects of the human form they describe:

- Shape features: Edge-based features are commonly used as the shape of an object can be properly captured by the edges. Examples include:
 - Parallel edge segments, rectangular contours
 - Binary contours followed by template matching, which is sensitive and fragile in images with cluttered background.
 - Edgelets, shapelet, adaptive contour features
 - Scale Invariant Feature Transform (SIFT) [P-9], Histogram of Oriented Gradients (HOG) [P-10], both of these will be described in more details below
- Appearance features: Image intensity or colour have been used to compute appearance features to compensate for the limitation of edge-based features. Examples include
 - Haar wavelets, generalized Haar
 - Local Binary Pattern (LBP), which is robust against illumination changes, discriminative power and computational simplicity
 - Local receptive fields, second order statistics of colours, i.e. colour self-similarities
- Motion features: For video, motion information can be exploited to improve the discriminative power by including temporal evolution.
 - Temporal differences: An important property of human movement is the periodicity, which can help distinguish between non-rigid objects (e.g. human) and rigid objects (e.g. cars). Moreover, gait models can be used to discriminate the human motions from other cyclic motions.
 - Optical flows: Histograms of flows (HOF) computed based on differential flows similar to the HOG can describe the boundary motions as well as internal/relative motions.
- Combining features: Various cues have been combined to improve detection performance since they can compensate each other. Examples of combination include:
 - Edge orientation histogram + rectangular features
 - Edgelet + HOG, Edgelet + HOG + covariance
 - HOG + LBP, HOG + HOF
 - HOG / HOF / LBP + second order statistics of colours

Object representation is about how an object is decomposed into a number of local regions on which feature extraction is performed. The two approaches are:

- Grid-based representation: For example, HOG features are computed on a dense grid of uniformly spaced blocks. However, the blocks cannot adequately capture the actual shape of the object, and irrelevant information at every location on the grid could corrupt the feature vectors. It is sensitive to object deformation and articulation.

- Interest points-based representation: Features are computed at corners, scale-invariant points or edge map. Advantages over the grid-based representation: compact object descriptor can be created, more appropriate for representing non-rigid objects with high articulation such as humans. A drawback is that interest points are detected locally and independently without considering the spatial constraints and so is sensitive to clutter.

An object of interest can be organized in a global or local structure. The global approach focuses on the whole object while local methods organize an object as a set of parts constituting the whole. The parts are not necessarily semantic body parts of a human object. Local approach has the advantage of being able to describe objects with high articulation and cope with occlusion, but it needs to validate the configurations of parts to form a meaningful object. For FSAR where the human target is of low-resolution, the parts approach is not applicable.

[P-8] reviews skeletonization which is a transformation of a component of a binary image into a subset of the original component. The motivation is to reduce the amount of data or to simplify the shape of an object in order to find features for recognition and classification. The three categories of algorithms include distance transform, critical points connection and iterative thinning. It is easier to understand by looking at some examples shown in Figure 2-2.

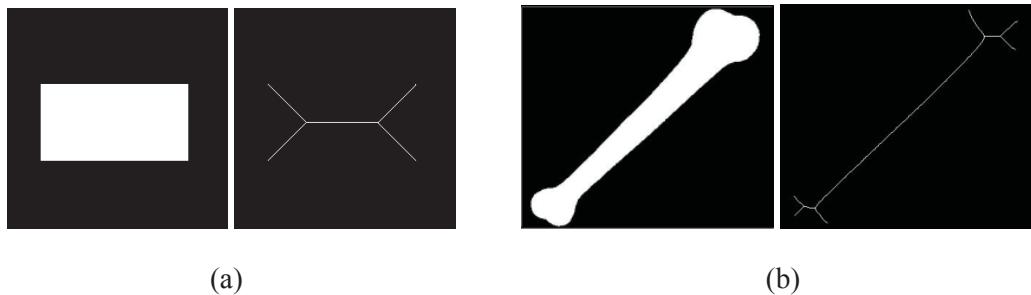


Figure 2-2 Skeleton Examples (a) Distance Transform (b) Iterative Thinning

Scale Invariant Feature Transform (SIFT) [P-9] was developed for image feature generation in object recognition applications. The features are invariant to image translation, scaling, rotation and partially invariant to illumination changes and affine or 3D projection. Interest point locations are defined as the maxima and minima of the difference of Gaussian applied in scale space, to ensure they are stable for matching. After locating the interest points, highly distinctive local descriptors are then computed for each of the features to facilitate recognition. Figure 2-3 shows an example of SIFT features matching across large baseline and viewpoint variation. It can be seen that most matches are correct, thanks to the invariance and discriminative nature of SIFT features. The size and orientation of the squares correspond to the scale and orientation of the SIFT features.

Being one of the earliest scale-invariant local distinctive features, SIFT (first published in 1999) has proven successful in a number of applications including object recognition, image stitching, mobile robot localization and mapping. SIFT has also inspired subsequent feature detectors such as SURF (Speeded Up Robust Features), GLOH (Gradient Location and Orientation Histogram), ORB (Oriented FAST and Rotated BRIEF), etc.



Figure 2-3 Example of Wide Baseline Matching Between Two Images With SIFT

Histogram of Oriented Gradients (HOG) [P-10] was first developed for pedestrian detection. The object image is normalized to 64x128 and uniformly divided into a dense grid of overlapping blocks. Each block was then split into 2x2 non-overlapping cells of size 8x8 pixels where HOGs were extracted. The object was encoded into a feature vector by concatenating the HOGs computed at each cells and blocks.

As HOG operates on localized cells, it maintains invariance to geometric and photometric transformation except for orientation. Moreover, the use of coarse spatial sampling, fine orientation sampling and strong local photometric normalization allows individual body movement of pedestrians to be ignored as long as they maintain an upright position. Although SIFT has a more compact representation, it was designed for specific objects, and was limited in the ability to generalize to object classes.

The FAST (Features from Accelerated Segment Test) feature detector was first proposed in [P-11] and further enhanced in [P-12]. To determine whether a pixel C is a FAST feature, a circle of 16 pixels surrounding C is examined, as illustrated in Figure 2-4. A feature is detected at C if the intensities of at least 12 contiguous pixels are all above or below the intensity of C by some threshold. This test can be optimized by only examining pixels 1, 9, 5 and 13, since a feature can exist only if at least 3 of them are all above or below the intensity of C by the threshold. The feature vector is the pixel intensities from the 16 pixels. A high performance tracking system was shown in [P-11] thanks to the high speed FAST corner detection.

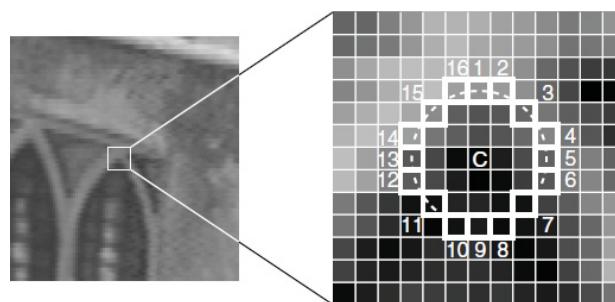


Figure 2-4 FAST Feature Detection Example (Source: [P-11])

[P-12] addresses several weakness of the original FAST detector by using a machine learning approach. During training, features are detected from a set of images (preferably from the target application domain) using a slow algorithm that tests all 16 locations around the pixel. A decision tree is created by recursively choosing a location that yields the most information about whether the candidate pixel is a corner. The decision tree is then converted into C-code, as a long string of nested if-then-else statements to use as a corner detector. The purpose is to learn from the training images what order to test the locations to minimize the number of tests. The results show that the learned FAST detector is faster than the original detector, as the average number of tests per pixel has decreased from 2.8 to 2.39. Apart from the significant speed up, test results also show that FAST exhibits high levels of repeatability in comparison with other feature detectors. However, FAST does not produce multi-scale features and is not very robust to the presence of noise.

[P-14] proposes ORB (Oriented FAST and Rotated BRIEF), building on the FAST feature detector and the Binary Robust Independent Elementary Features (BRIEF) descriptor, both of which have good performance and low cost. ORB addresses their limitations, in particular the lack of rotational and scale invariance. FAST does not include an orientation operator nor a measure of cornerness, which are addressed by ORB. Experiments show that ORB is two orders of magnitude faster than SIFT, while performing as well in many situations. ORB runs at 7 Hz (640×480 resolution) on a cellphone with 1 GHz ARM chip, and hence is suitable for real-time feature tracking on embedded devices.

[P-13] presents an approach for measuring similarity between images/videos, based on the internal layout of local self-similarities. The Local Self-Similarity (LSS) descriptor is measured densely throughout the image at multiple scales, while accounting for local and global geometric distortions. Figure 2-5 shows an example, where a 5×5 image patch centered at pixel q is correlated with a larger surrounding image region of radius 40 pixels, resulting in a local internal correlation surface. The correlation surface is then transformed into a log-polar representation, with 80 bins (20 angles, 4 radial intervals). The maximal values in those bins form the 80 entries of the descriptor vector, which is then normalized. By matching an ensemble of LSS descriptor vectors, it can detect objects in cluttered images well, even with rough hand-sketches, as shown in Figure 2-6.

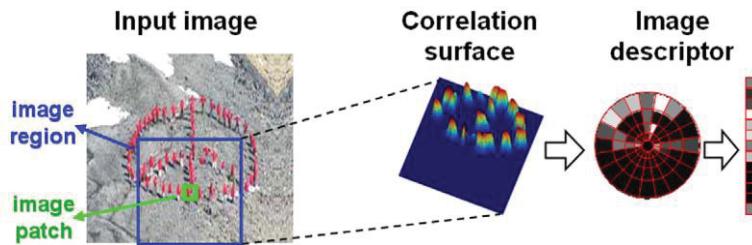


Figure 2-5 Local Self-Similarity Descriptor Example at an Image Pixel (Source: [P-13])



Figure 2-6 Detection of a Person Raising Both Arms (a) A Hand-sketched Template (b) Detected Locations in Other Images (Source: [P-13])

[P-15] proposes two low-dimensional variants of the LSS using Principal Component Analysis (PCA), to improve the invariance and performance of the original LSS descriptor.

- PCA-LSS: By replacing the gradient magnitude feature used in PCA-SIFT with the LSS feature, followed by PCA, the resulting PCA-LSS descriptor is smaller (36 dimensions) than the original LSS vector (80 dimensions).
- PLSS: The top 36 eigenvectors of LSS descriptors extracted from a training image set are pre-computed offline, which capture the major variances of the LSS descriptors. A low-dimensional variant of LSS using PCA (PLSS) is obtained by applying PCA with the eigenspace formed by these eigenvectors, resulting in 36 dimensions.

Experiments were performed to compare the proposed descriptors with other existing descriptors, showing that the PLSS descriptor outperforms the original LSS and also SIFT.

Table 2-7 compares these feature detectors in terms of the FSAR criteria. All of them are available in Open Source Computer Vision Library (OpenCV).

Table 2-7 Comparison of Selected Papers on Feature Extraction

Feature	Input Data	Feature Type	Computational Speed	Human Detection at Long Range?	Suitable Applications	Notes
Skeleton	Binary images	Global	Fast	Yes	<ul style="list-style-type: none"> • Shape recognition 	
SIFT	Images	Local, sparse	Slow (GPU version available)	<ul style="list-style-type: none"> • May not generalize to human detection 	<ul style="list-style-type: none"> • Object recognition • Image matching 	Not free for commercial use
HOG	Images	Local, dense	Slow (GPU version available)	Yes	<ul style="list-style-type: none"> • Object detection • Human detection 	Require SVM with offline training

Feature	Input Data	Feature Type	Computational Speed	Human Detection at Long Range?	Suitable Applications	Notes
FAST	Images	Local, sparse	Very fast	<ul style="list-style-type: none"> Not scale-invariant May not generalize to human detection 	<ul style="list-style-type: none"> Tracking Image matching 	
ORB	Images	Local, sparse	Very fast	<ul style="list-style-type: none"> May not generalize to human detection 	<ul style="list-style-type: none"> Tracking Image matching Object recognition 	
LSS	Images, video	Local, dense	Fast	Yes	<ul style="list-style-type: none"> Object detection Image matching Action detection in video 	

2.2.4 Classification/Recognition

The term classification is often used for coarse categorization, while the term identification is used for fine categorization, and they are sometimes used synonymously with the term recognition. For this report, both classification and recognition are regarded as discerning a type of object, e.g. vehicles versus human, while identification distinguishes which model of the vehicles or which particular person. For the FSAR application, we are interested in techniques that can classify the foreground objects as either human or clutter. Table 2-8 lists the selected papers being reviewed.

In [P-10], HOG descriptors are passed to some recognition system based on supervised learning. The Support Vector Machine (SVM) classifier is a binary classifier which looks for an optimal hyperplane as a decision function. It is first trained with many known positive and negative examples, after which the SVM classifier can make decisions regarding the presence of human in additional test images.

Table 2-8 Selected Papers on Classification/Recognition

#	Paper Title	Authors	Source	Year
P-17	Multimodal Approach to Human-Face Detection and Tracking	P. Vadakkepat, P. Lim, L.C. De Silva, L. Jing, L.L. Ling	IEEE Transactions on Industrial Electronics, 55(3)	2008
P-18	Development of an Infrared Imaging Classifier for UGS	B. D'Agostino, M. McCormack, B. Steadman	SPIE Vol. 7693	2010
P-19	Dismounted Human Detection at Long Ranges	A.E. Bell	SPIE Vol. 8049	2011
P-20	Human Detection and Tracking under Complex Activities	B. Cancela, M. Ortega, M.G. Penedo	International Conference on Computer Vision Theory and Applications	2013

Figure 2-7 (a) shows the average HOG gradient image over the training examples. Figure 2-7 (b) and (c) show the maximum positive and negative SVM weights respectively, indicating the emphasis on the silhouette contours, especially the head, shoulders and feet. Figure 2-7 (d) shows a test image and Figure 2-7 (e) shows the computed HOG descriptor. Figure 2-7 (f) and (g) show the HOG descriptor weighted by the positive and the negative SVM weights respectively.

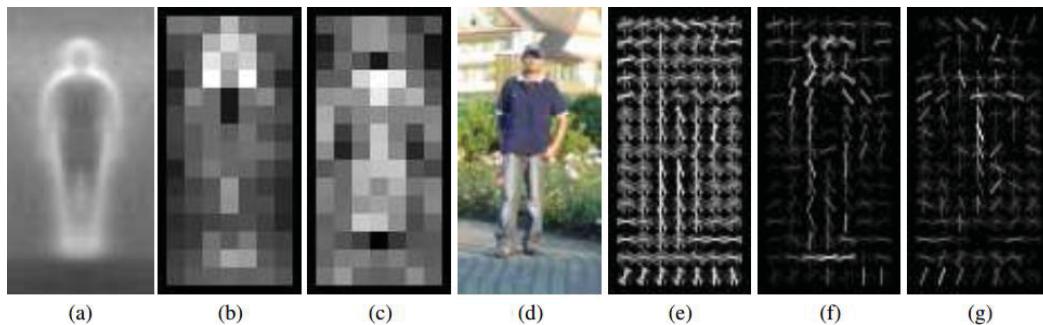


Figure 2-7 HOG Cue Mainly on Silhouette Contours (Head, Shoulders & Feet) (Source: [P-10])



Figure 2-8 Test Image at Four Scales (a) Original Size (b) Factor-2 Reduction (c) Factor-4 Reduction (d) Factor-8 Reduction (Source: [P-19])

[P-19] describes a recent study to evaluate how the HOG with SVM approach for human detection performs at long ranges. The results show that HOG remains effective even at long distances, for example, the miss rate and false alarm rate are both only 5% for humans that are 12 pixels high and 4-5 pixels wide. Figure 2-8 shows an example test image at four scales, in which the person is only 12 pixels high in the case of factor-8 reduction, corresponding to a Ground Sample Distance (GSD) of 15cm/pixel.

Only single images are used currently without any image enhancement. As potential future work, the paper suggests HOG + SVM coupled with other techniques like super-resolution or methods that exploit video data, could provide human detection at extremely long ranges. This finding is very applicable to FSAR, where the standoff distance is up to 600m and the size of a person could be as small as 10×10 to 16×16 pixels.

[P-17] focuses on the interaction capability of mobile robots, particularly in detecting, tracking and following human subjects. Rather than human detection, the system performs face detection based on the skin color in the UV color space. A neural network is used to learn the

skin and non-skin colors. Once the face is detected, the tracking algorithm is activated where information from sonar and tactile sensors is also utilized.

[P-18] presents a target classifier using IR imagery to recognize human, vehicle, animal and clutter for an unattended ground sensor. The purpose is to transmit only images that contain potential targets of interest to minimize bandwidth resources. The system includes a number of features that can be selected, such as height/width ratio, peak/clutter ratio, area feature, gradient mean ratio, etc. The optimal set of features can be evaluated experimentally to provide the best performance, using various classifiers such as Bayesian, neural networks, etc.

[P-20] proposes a new methodology for detecting and tracking people under uncontrolled and complex scenarios. A background subtraction technique is first used to detect moving pixels in the scene. The Viola-Jones classifier is used to detect every possible human being in the scene, to be confirmed by the HOG with SVM classifier. Viola-Jones classifier is fast but lacks classification accuracy, while HOG with SVM is a better classifier but takes more time. The use of two classifiers helps improve the processing speed.

Table 2-9 compares these papers in terms of the FSAR criteria.

Table 2-9 Comparison of Selected Papers on Classification/Recognition

Paper	Input Data	Technique	Low Resolution/ Long Range?	Sensor	Targets Detected	Real-time Processing?
[P-17]	Video from moving camera	Neural network for skin-color model	No	Visible	Face of human	NA
[P-18]	Video from fixed camera	Various features with classifier	Yes (up to 150m)	IR	Human, vehicle, animal	NA
[P-19]	Images	HOG with SVM	Yes (person of 12 pixels high)	Visible	Human	NA
[P-20]	Video from fixed camera	Viola-Jones classifier + HOG with SVM	Yes	Visible	Human	No (4 Hz for 640x368 resolution on Pentium Quadcore 2.4 GHz)

2.2.5 Object Tracking

In the FSAR application, since both the rifle-mounted video camera and the human target could be moving, it is necessary to keep track of the target over frames once it has been recognized. This would be more efficient than trying to detect and recognize the target again at each frame and would also help reduce false alarms using the temporal information.

As the object tracking literature is vast, the papers reviewed here, as listed in Table 2-10, focus on techniques suitable for human tracking. Most of the tracking systems include some predictors to predict where the target will be. Such prediction would be useful for the FSAR application to handle latency between the target recognition and the shot placement.

Table 2-10 Selected Papers on Object Tracking

#	Paper Title	Authors	Source	Year
P-21	Object Tracking: A Survey	A. Yilmaz, O. Javed, M. Shah	ACM Computing Surveys, 38(4)	2006
P-22	Infrared Human Tracking with Improved Mean Shift Algorithm based on Multicue Fusion	X. Wang, L. Liu, Z. Tang	Applied Optics, 48(21)	2009
P-23	Robust and Efficient Fragments-based Tracking using Mean Shift	F.L. Wang, S.Y. Yu, J. Yang	International Journal of Electronics and Communications, 64(7)	2010
P-24	Multiple Target Tracking Using Cognitive Data Association of Spatiotemporal Prediction and Visual Similarity	Y.M. Seong, H. Park	Pattern Recognition, 45(9)	2012
P-25	A Multiple Targets Appearance Tracker Based on Object Interaction Models	G.R. Li, W. Qu, Q.M. Huang	IEEE Transactions on Circuits and Systems for Video Technology, 22(3)	2012
P-26	Object Tracking with Adaptive HOG Detector and Adaptive Rao-Blackwellised Particle Filter	S. Rosa, M. Paleari, P. Ariano, B. Bona	SPIE Vol. 8301	2012
P-27	Human Tracking with an Infrared Camera using a Curve Matching Framework	S.J. Lee, G. Shah, A.A. Bhattacharya, Y. Motai	Eurasip Journal on Advances in Signal Processing, 99	2012

[P-21] is a survey article on object tracking, including taxonomy of tracking methods as shown in Figure 2-9. The three categories are:

- Point Tracking: Tracking is formulated as the correspondence of detected objects represented by points across frames. The correspondence could be complicated especially in the presence of occlusion, misdetection, appearance, disappearance of objects. It can be further divided into deterministic and probabilistic approaches which include Kalman filter and particle filter.
- Kernel Tracking: Kernel refers to the object shape and appearance, such as a rectangular template, an elliptical shape with an associated histogram. Tracking is performed by computing the motion of the primitive object region from one frame to the next. It can be further divided into multi-view based and template based approaches which include mean shift.
- Silhouette Tracking: Tracking is performed by estimating the object region in each frame, based on an object model generated using the previous frames. This can handle complex shapes that cannot be well described by simple geometric shapes. It can be further divided into shape matching and contour evolution.

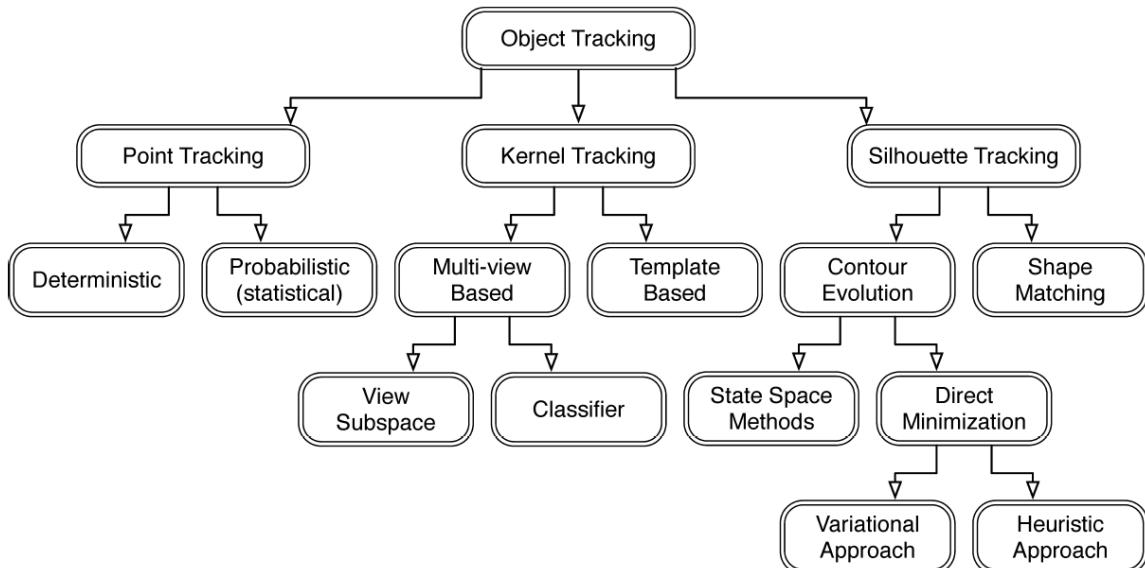


Figure 2-9 Taxonomy of Tracking Methods (Source: [P-21])

[P-22] presents an improved mean shift algorithm for tracking infrared human targets using multicue fusion. Mean shift tracking is a deterministic approach which is less computationally intensive than particle filter. Many existing mean shift-based algorithm rely on a single cue and cannot cope with complex background clutter. Therefore, multicue (gray and edge cues) fusion is used, and motion-guided cues are proposed as both the gray and edge cues become useless under partial/complete occlusion.

[P-23] describes human fragments-based tracking using mean shift by dividing the target into multiple distinctive fragments. The human fragments are extracted based on a graph cut technique after the user marks the target region. The use of multiple fragments helps maintain the spatial information while each fragment is weighted to account for partial occlusion. Experimental results show that the performance of the proposed algorithm is better than the basic mean shift algorithm, during partial occlusion or long-time occlusion.

For multiple target tracking, [P-24] proposes a data association process based on two primary components of visual features and spatiotemporal prediction. Moreover, the change perception and visual distinguishability are used to adaptively combine the two primary components. The prediction is filtered by the change perception mask to remove clutters. The proposed system shows consistent tracking performance on video sequences containing small targets with low visual distinguishability and irregular motions.

Instead of tracking each target independently, [P-25] augments a kernel-based tracker with object interaction models because a moving object's motion could also be impacted by other neighboring objects. For human and vehicle tracking, the object usually moves toward a particular direction but detours when close to others to avoid collision. By defining virtual destination of a target and virtual gravity to indicate its attraction force, a new cost function can be embedded into the kernel-based tracker. Experimental results show better tracking performance is obtained with the object interaction model.

[P-26] presents a system for real-time visual human tracking for mobile robots, to facilitate human-robot interaction for future planetary exploration scenarios. HOG with SVM are used as the human detector, followed by an adaptive Rao-Blackwellised particle filter to track the detected human. An advantage of particle filter over classic Kalman filter is the ability to cope with non-linearity and non-Gaussianity, which are critical in the case of moving objects. Moreover, multi-model distributions can be modeled by particle filter.

[P-27] proposes a tracking algorithm that combines a curve matching framework and Kalman filter to enhance the prediction accuracy of human tracking. Human target often have a prominent moving pattern such as a cyclic pattern, which is not captured by Kalman filter alone. Curve matching compares the current motion with the motion trajectory history, which allows the algorithm to predict the next human movement better. Experimental results show that a mobile robot can track a human better with the proposed algorithm.

Table 2-11 compares these papers (except the survey paper) in terms of the FSAR criteria.

Table 2-11 Comparison of Selected Papers on Object Tracking

Paper	Input Data	Technique	Low Resolution/ Long Range?	Sensor	Targets Tracked	Real-time Processing?	Implement- ation
[P-22]	Video from fixed camera	Mean shift + multiple cues (gray & edge)	Yes	IR	Human	Yes (19 Hz for 160x120 resolution on Celeron 2.8 GHz)	Matlab
[P-23]	Video from fixed camera	Mean shift + multiple fragments	No	Visible	Human	Yes (>30 Hz on Pentium IV 2.66 GHz)	C++
[P-24]	Video from fixed camera	Visual similarity + spatiotemporal prediction + change perception + visual distinguishability +	Yes	Visible	Human & vehicles	Yes (6 – 13 Hz for 768x576 to 384x288 resolution on Pentium 2.67 GHz)	C++
[P-25]	Video from fixed camera	Kernel-based + object interaction model	Yes	Visible	Human & vehicles	Yes (>20 Hz on Pentium IV 3.4 GHz)	C++
[P-26]	Video from moving camera	HOG + SVM + particle filter	No	Visible	Human	Yes (20 Hz for 320x240 resolution on 2.4 GHz CPU)	C++ with OpenCV
[P-27]	Video from moving camera	Kalman filter + curve matching	No	IR	Human	Yes (15 Hz for 640x480 resolution)	NA

2.3 End-to-End ATC Systems

While the papers in the previous section focus on the individual steps of ATC, the papers reviewed in this section described end-to-end ATC systems. We divide the papers into ATC systems for infrared spectrum, visible spectrum and multi-sensor systems. Table 2-12 and Table 2-13 list the selected papers on ATC systems for infrared spectrum and visible spectrum respectively, while Table 2-14 lists the ones on multi-sensor ATC systems.

Table 2-12 Selected Papers on ATC Systems for Infrared Spectrum

#	Paper Title	Authors	Source	Year
P-28	Development of a Portable Infrared Surveillance System with Auto Target Cueing Capability	P. Laou, J. Maheux, Y. de Villers, J. Cruickshank, D. St-Germain, T. Rea, J. Cote, P. Vallee, A. Morin, P.O. Belzile, V. Labbe, N. Bedard-Maltais	Army Science Conference, Orlando	2008
P-29	Low False Alarm Target Detection and Tracking within Strong Clutters in Outdoor Infrared Videos	C. Li, J. Si, G.P. Abousleman	Optical Engineering, 49(8)	2010

Table 2-13 Selected Papers on ATC Systems for Visible Spectrum

#	Paper Title	Authors	Source	Year
P-30	Evaluation of USC Human Tracking System for Surveillance Videos	B. Wu, X. Song, V.K. Singh, R. Nevatia	International Evaluation Workshop on Classification of Events, Activities and Relationships, (CLEAR)	2006
P-31	Real-Time Human Detection, Tracking and Verification in Uncontrolled Camera Motion Environments	M. Hussein, W. Abd-Almageed, Y. Ran, L. Davis	IEEE International Conference on Computer Vision Systems (ICVS)	2006
P-32	Probabilistic People Tracking with Appearance Models and Occlusion Classification: the AD-HOC System	R. Vezzani, C. Grana, R. Cucchiara	Pattern Recognition Letters, 32(6)	2011

Table 2-14 Selected Papers on Multi-Sensor ATC Systems

#	Paper Title	Authors	Source	Year
P-33	Active People Recognition using Thermal and Grey Images on a Mobile Security Robot	A. Treptow, G. Cielniak, T. Duckett	IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)	2005
P-34	Shape and Motion-based Pedestrian Detection in Infrared Images: a Multi Sensor Approach	B. Fardi, U. Schuenert, G. Wanielik	IEEE Intelligent Vehicles Symposium	2005

2.3.1 Infrared ATC Systems

[P-28] presents a portable real-time system for longwave infrared (LWIR) surveillance with auto target cueing capabilities, which is part of the Advanced Linked Extended Reconnaissance Targeting (ALERT) Technical Demonstration Project (TDP). The system is equipped with two LWIR channels, one for wide field-of-view and one for narrow field-of-view. It processes LWIR video at 30 Hz at 320×240 resolution. The ATC processing steps are as follows:

1. Enhance image with an auto-adjust contrast algorithm
2. Compute background image estimate and subtract from registered images
3. Apply adaptive-threshold image binarization method to find potential moving objects
4. Apply morphological filter to remove irrelevant objects and agglomerate adjacent blobs
5. Compute blob characteristics and track objects over time based on correlation
6. Extract image chips and classify into three categories (vehicle, human and clutter)
7. Send vehicle and human positions to display processing unit

[P-29] presents a surveillance system to detect and track vehicles in motion and people in transit using a stationary IR camera. The system demonstrates very few false alarms, high detection accuracy and consistent tracking with real-world IR video of complex background and motion clutter, as well as small and blurred moving targets. The target motion pattern is examined based on the fact that an independent moving target (a person or a vehicle) usually follows a smooth trajectory within a short time window, whereas a false alarm tends to appear randomly. During walking-person recognition, the shape of the human figure is chosen as the key feature, followed by the SVM which has proven to be a robust supervised classifier.

Table 2-15 compares these systems in terms of the FSAR criteria.

Table 2-15 Comparison of ATC Systems for Infrared Spectrum

Paper	Input Data	Low Resolution/ Long Range?	Sensor	Targets Detected	Real-time Processing?	Implementation
[P-28]	Video from moving camera	Yes	IR	Vehicle and human	Yes. (30 Hz for 320x240 resolution)	NA
[P-29]	Video from fixed camera	Yes	IR	Vehicles and walking person	Yes. (10 Hz for 320x240 resolution)	OpenCV and C++

2.3.2 Visible ATC Systems

[P-30] presents a human tracking system for surveillance video and the evaluation results. The key steps of the system are as follows:

1. Detect motion by comparing pixel colour to an adaptively learned background model
2. Search for humans in the moving blobs only (this prevents false alarms on static scene objects and also omits static persons in the scene)

3. Combine shape-based tracking with motion-based blob tracking to increase accuracy
4. Verify by 3D speed to discriminate humans from vehicles (requires calibration parameters)

[P-31] describes a real-time system for human detection from a freely moving platform. The focus of the paper is on system robustness and efficiency rather than on the algorithmic issues. Robustness is achieved through integration of algorithms, for human detection, tracking & motion analysis, in one framework so that the final decision is based on the agreement of more than one algorithm. Efficiency is achieved through multi-threaded design and usage of a high performance computer vision & image processing library. Different visual cues are used to filter out false alarms:

1. Human detection algorithm uses the shape cue to decide whether part of the image contains a human
2. Tracking algorithm uses the intensity cue to track the object over time
3. Motion analysis algorithm uses the motion periodicity cue to verify the object moves like a human

[P-32] presents a framework for multiple people tracking in video surveillance applications with large occlusions. The key contribution was on overcoming large and long-lasting occlusions by using an appearance driven tracking model. It models non-visible region which is classified into three classes: dynamic occlusions, scene occlusions and apparent occlusions. It assumes that different objects may be distinguished by their colour. The size of the human objects is fairly large in the experimental results shown.

Table 2-16 compares these systems in terms of the FSAR criteria.

Table 2-16 Comparison of ATC Systems for Visible Spectrum

Paper	Input Data	Low Resolution/ Long Range?	Sensor	Targets Detected	Real-time Processing?	Implementation
[P-30]	Video from fixed camera	Yes	Visible	Human	No (5 seconds per frame on 2.8GHz Pentium CPU)	C++ with OpenCV
[P-31]	Video from moving camera	Yes	Visible	Human	Yes (15 Hz)	Intel Integrated Performance Primitives (IPP) library, OpenThreads library
[P-32]	Video from fixed camera	No	Visible	Human	Yes (10 Hz)	NA

2.3.3 Multi-Sensor ATC Systems

[P-33] presents a vision-based approach to detect, track and identify people from a mobile robot in real time. Thermal imagery is first used to detect person from a larger distance. Then, the robot would drive towards the person while tracking with a particle filter technique. When the robot is close by, it uses grayscale images from its pan-tilt camera to track the face, which is fed into a recognition system to identify the person. In this case, the two sensors are used sequentially to complement each other, as the thermal camera is better in locating people from a distance and the visible camera is necessary for face tracking & face recognition.

[P-34] describes a multi-sensor approach to detect and track pedestrians using shape and motion cues. A Kalman filter is used to fuse the infrared image processing output with the laser scanner processing output, to synchronize sensor data from non-synchronized sources. The laser scanner helps to ensure further processing is restricted on regions of interest only. Shape extraction is used in which the extracted contour is compared with reference sets in the Fourier domain and the cyclical shape of motion is used to recognize people. The size of people is fairly large in the experimental results shown.

Table 2-17 compares these systems in terms of the FSAR criteria.

Table 2-17 Comparison of Multi-Sensor ATC Systems

Paper	Input Data	Low Resolution/ Long Range?	Sensor	Targets Detected	Real-time Processing?	Implementation
[P-33]	Video from moving camera	Yes	Thermal & visible	Human	Yes (80 Hz on Athlon XP 1600)	NA
[P-34]	Video from moving camera	No	IR & laser scanner	Human		NA

2.4 ATC Performance Evaluation

Performance evaluation is essential for ATC system development, as it is important to compare the results with ground truth and other ATC systems. Table 2-18 lists the selected papers on ATC system performance evaluation.

Table 2-18 Selected Papers on ATC System Performance Evaluation

#	Paper Title	Authors	Source	Year
P-35	START for Evaluation of Target Detection and Tracking	J. Irvine, S.K. Ralph, M.R. Stevens, J. Marvel, M. Snorrason, D. Gwilt	SPIE Vol. 5807	2005
P-36	Infrared Sensor Modeling for Human Activity Discrimination Tasks in Urban and Maritime Environments	D.M. Deaver, E. Flug, E. Boettcher, S.R. Smith, B. Miller	Applied Optics, 48(19)	2009

#	Paper Title	Authors	Source	Year
P-37	Human Target Acquisition Performance	B.P. Teaney, T.W. Du Bosq, J.P. Reynolds, R. Thompson, S. Aghera, S.K. Moyer, E. Flug, R. Espinola, J. Hixson	SPIE Vol. 8355	2012

[P-35] presents a truthing system called Scoring, Truthing, And Registration Toolkit (START) with two components: a truthing component that assists in the automated construction of ground truth, and a scoring component that assesses the performance of a given algorithm relative to the ground truth. This system is specifically for video datasets, as it is very tedious and error-prone to label a large number of video frames. After the user manually marks the target in the first frame, it will automatically track the target chip in the subsequent frames. Such tool could be very useful for ATC performance evaluation in FSAR as it can help produce ground truth data for the large amount of collected video data.

[P-36] and [P-37] are related to evaluation of human observer performance rather than ATC, i.e. how well can a human observer identify the activity based on the sensor data. US Army's NVThermIP model is a standard sensor performance model that estimates target acquisition performance based on both sensor design parameters and measured calibration factors. It uses a metric called Targeting Task Performance (TTP) to compare resolution and sensitivity provided by a given sensor system. Historically, it is calibrated by presenting static imagery to observers and measuring average probabilities of recognizing the targets.

The task of human activity discrimination in video data presents new challenges, as it involves dynamic scene where motion cues are essential. [P-36] discusses the challenges in representing the human activity task, establishment of new processing methods and new standards for defining simple target metrics. Both the Johnson and TTP metrics were analyzed, showing that the Johnson method provides better model for static image data while the TTP metric performs better for the dynamic scene data.

Since the battlefield has now shifted from armored vehicles to armed insurgents, target acquisition performance involving humans as targets is vital for modern warfare. [P-37] described the experiments conducted by the US Army involving human targets: human activity, weapon/non-weapon, and two-hand object identification. Some example thermal images from the experiments are shown in Figure 2-10 and Figure 2-11. Such dataset is also applicable to FSAR to evaluate ATC performance. The paper defines a set of standard task difficulty values for identification and recognition associated with human target acquisition performance. One of the findings indicates that motion cues from video data heavily influenced the ability of the observer to identify a particular action from the set.

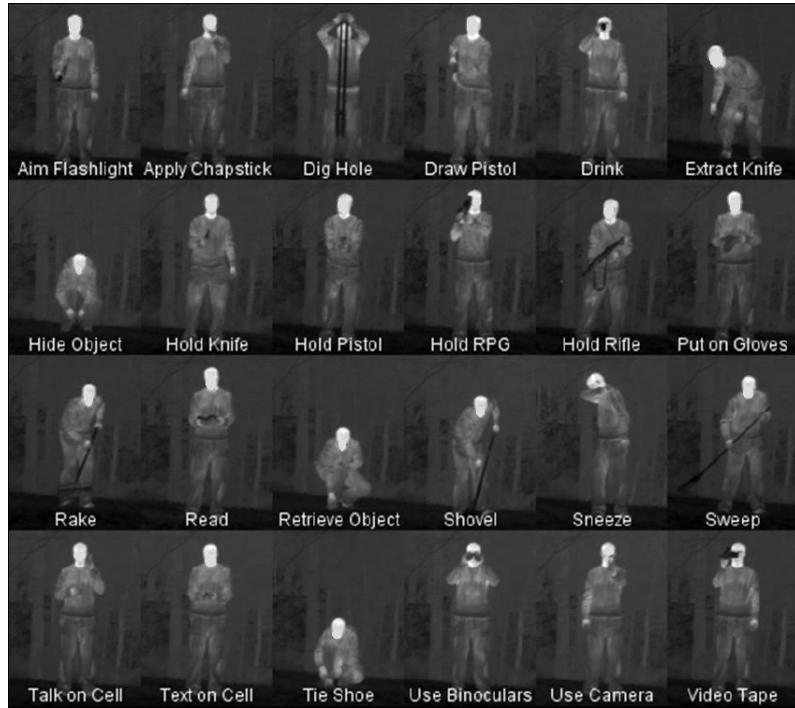


Figure 2-10 Thermal Images of Human Activities (Source: [P-37])



Figure 2-11 Thermal Images of Two-Hand Object Identification (Source: [P-37])

It is highly desirable to have test data with ground truth under wide range of conditions for testing. A number of video surveillance datasets with ground truth have been made available, such as Video and Image Retrieval and Analysis Tool (VIRAT) datasets [R-1], Performance Evaluation of Tracking and Surveillance (PETS) datasets [R-2], which greatly facilitate comparison of different algorithms. More specifically, a number of human object datasets have been made available for the evaluation of human detection algorithms over the last decade. These datasets are collected from different scenarios and can be used as benchmarking for various applications of human detection [P-7]:

- General purpose person detection algorithms for image retrieval
 - MIT, INRIA, Penn-Fudan, USC-A, USC-C datasets

- Surveillance application
 - USC-B and CAVIAR datasets
- Pedestrian detection in driving assistance systems
 - Caltech, TUD, CVC, DaimlerChrysler, ETH datasets

They are most probably not suitable for FSAR application where the human is far away and the camera is moving. US Army's Night Vision and Electronic Sensors Directorate (NVESD) has been evaluating various algorithms since the early 1980s but the data and ATC system details only appeared in the classified literature. United States (US) Department Of Defense (DOD) has recently made available over 300 GB of imagery data (IR and visible) of tactical vehicles, civilian vehicles and people in realistic tactical scenes with ground-truth [P-2], which could be useful for FSAR.

2.5 COTS ATC Software/SDK

This Section reviews Commercial Off-The-Shelf (COTS) software or Software Development Kit (SDK) applicable to ATC or the various processing steps. COTS software refers to software packages that can be used out of the box to perform ATC or some of the processing steps, while SDK refers to libraries that offer functionalities for ATC software development. The information is based on publicly available information on vendor websites, which typically do not provide the pricing information.

2.5.1 COTS Software

For the FSAR application, we are mainly interested in software that can handle mobile cameras since the rifle-mounted video camera would be moving. However, a number of COTS software for fixed camera is also included, as some pre-processing step may potentially be performed to stabilize the video beforehand.

There are many companies offering COTS intelligent video surveillance systems for fixed camera surveillance, driven by the large number of Closed Circuit Television (CCTV) in use nowadays. On the other hand, there is relatively few COTS software for mobile cameras, as it is harder than fixed camera surveillance and has lower demand. The software needs to handle the camera motion to find moving objects. These software packages are often designed for processing aerial video feeds from Unmanned Aerial Vehicles (UAV) or manned airborne platforms.

Table 2-19 compares the various COTS software packages in terms of FSAR criteria.

2.5.2 SDK Applicable to ATC Development

There are a number of SDKs that address some of the ATC processing steps. Such functionalities can be integrated with ATC system development to avoid re-inventing the wheels and may be applicable to Tasks 5 to 7 of this project.

Table 2-20 compares the various SDKs in terms of FSAR criteria.

2.6 Concluding Remarks

The literature review presented above has indicated various promising approaches for human target detection applicable to FSAR, addressing moving camera, long range and low resolution requirements. However, most of the papers do not tackle the issue of false alarms in cluttered environment, which is the key challenge for ATC deployment in military operations.

The proposed plan for way forward:

- While the current report covers a wide range of topics at a high level, a more detailed but focused literature review can be conducted based on DRDC feedback, as a second iteration of this task.
- Evaluate some promising COTS packages with representative FSAR data to assess whether the performance of any COTS package is sufficient for operational use. This could be done using evaluation software provided by vendors, or software purchased by the project.
- Develop a new ATC system based on the promising approaches described, making use of suitable SDKs identified. A comprehensive evaluation with representative FSAR data is necessary to assess the performance, in particular for cluttered environments.

Table 2-19 Comparison of COTS Software Applicable to ATC

Company	COTS Software	Web-page	Key Capabilities	Input Data	Low resolution	Sensor	Real-time processing	Notes	Price
Sentient	Kestrel	[R-3]	Automatically detect moving objects in video from manned and unmanned aerial vehicles. <ul style="list-style-type: none">• Detect and cue operators to small moving targets (down to a few pixels) by indicating them on viewing screen.• Only detect any moving pixels, but does not distinguish among vehicles, people, or other moving objects.• Proven in theatre on various aerial platforms including Heron, ScanEagle & Shadow.	Video from moving camera	Yes	Visible and IR	Yes	NDA is needed to further discussion.	
Deep Vision	Person tracking & targeting	[R-4]	Provide software and hardware solutions for real-time intelligent machine perception capability such as <ul style="list-style-type: none">• Object tracking independent of sensor modality, sensor motion and object motion• Object recognition• Automatic target detection• Runtime target designation	Video from moving camera	Yes	Visible and IR	Yes	Vendor does not offer evaluation software, but can help processing some test data. NDA is needed for further discussion.	
iOmniscient	iQ series	[R-5]	Suite of detection products for surveillance video, such as <ul style="list-style-type: none">• Intrusion detection, perimeter protection, person counting, human behaviour analysis• Detection of remaining objects, removed objects, vandalism in crowded area	Video from fixed camera	Yes	Visible and IR	Yes	Vendor has confirmed that their products cannot handle video from moving camera	

Company	COTS Software	Web-page	Key Capabilities	Input Data	Low resolution	Sensor	Real-time processing	Notes	Price
Cognika	Pegasus	[R-6]	Designed for video feeds from moving camera such as aerial assets, vehicle mounted camera <ul style="list-style-type: none"> • Detect objects, their motion trajectories and activities in real-time • Apply stabilization, mosaicking and tracking to detect objects, events and activities • Visual search and monitoring technologies to locate when a specific type of vehicle or person appears 	Video from moving camera	Yes	Visible and IR	Yes		
Cognika	Perseus	[R-7]	<ul style="list-style-type: none"> • Fine-grained activity detection for very accurate surveillance • Classify vehicles, humans, animals • Alert on configured rules & set thresholds 	Video from fixed camera	Yes	Visible and IR	Yes		
Fraunhofer IOSB	ABUL	[R-8]	Designed for assisting video exploitation process and to disburden operator during critical mission, providing capabilities for surveillance & tactical reconnaissance <ul style="list-style-type: none"> • Video-based moving target indication • Tracking of interactively marked vehicles • Deployed in Afghanistan by German Forces 	Video from moving camera	Yes	Visible and IR	Yes		
HP Surveillance	Intelligence Scene Analysis System	[R-9]	Provide video analytics for identification of items, objects and events within video, assist users of CCTV system in the detection of activities. Functions include: <ul style="list-style-type: none"> • Object classification and tracking • Video motion detection • Behaviour analysis 	Video from fixed camera	Yes	Visible	Yes	Technology was developed by Autonomy, which was acquired by HP in 2011.	
ObjectVideo	OV6 Analytics Service	[R-10]	Offer field-proven software for video analytics and surveillance, including <ul style="list-style-type: none"> • Object detection, classification, tracking • Object removal/insertion, loitering, rendezvous 	Video from fixed camera	Yes	Visible and IR	Yes		

Company	COTS Software	Web-page	Key Capabilities	Input Data	Low resolution	Sensor	Real-time processing	Notes	Price
Eptascape	EptaAnalytics	[R-11]	Distributed software suite for performing behavioral analysis, automatic events detection and recording, based on video annotation created by MPEG-7 encoder. Examples include: <ul style="list-style-type: none">• Detection of unattended objects, parked cars• People tracking in shopping mall• Cars line crossing detection	Video from fixed camera	Yes	Visible and IR	Yes		
Aimetis	Symphony and VE Series Video Analytics	[R-12]	Offer a video platform for video management, video analytics, system integration & alarm management. Enterprise version includes VE Series Video Analytics, a suite of video surveillance tools including: <ul style="list-style-type: none">• Video motion detection, Motion tracking• Object classification, Left/removed item detection	Video from fixed camera	Yes	Visible and IR	Yes	Offers free 30 days trial	Symphony Enterprise at \$825
Honeywell	Active Alert Video Analytics	[R-13]	Provide video analytics to automatically <ul style="list-style-type: none">• Detect, analyze, track and classify behaviours of people and vehicles• Real-time scene analysis and alarms based on user-definable rules	Video from fixed camera	Yes	Visible and IR	Yes		
IntelliVision	Intelligent Video Analytics	[R-14]	Provide recognition suite with capabilities including: <ul style="list-style-type: none">• Recognition of faces and objects• Detection of motion, intrusion, object left, loitering• Video search and summary	Video from fixed camera	Yes	Visible	Yes		
MotionDSP	Ikena ISR	[R-15]	Software product for processing, exploitation and dissemination, with capabilities including: <ul style="list-style-type: none">• Real-time enhancement• Image stabilization, super-resolution• Visual moving target indicator	Video from moving camera	Yes	Visible and IR	Yes	Leverages off-the-shelf GPUs to achieve advanced video processing at real-time	

Table 2-20 Comparison of SDK Applicable to ATC Development

Company	COTS Software	Web-page	Key Capabilities	Supported Platforms	Languages / Programming Environments	Notes	Price
OpenCV	OpenCV	[R-15]	Open source computer vision library for real-time computer vision, with more than 2500 optimized algorithms including state-of-the-art computer vision and machine learning algorithms • Detect and recognize faces, identify objects, classify human actions, track moving objects, etc.	• Windows • Linux/Mac • Android • iOS • NVIDIA GPU	• C/C++ • Wrappers for C#, Python, Java	Very widely used by companies, university research groups and government bodies	Free
Charles River Analytics	VisionKit 1.0	[R-17]	A tool suite for developing real-time computer vision applications. Provide building blocks such as • Find targets, automate surveillance, recognize motion • Enhance, segment and classify 2D images and 3D data	• Windows	• C/C++	A case study describes the use of the VisionKit to develop ATR software rapidly	\$499
2D3	Tungsten Media Toolkit	[R-18]	Provide comprehensive modules to develop real-time digital media solutions and computer vision applications • Moving target detection • Super resolution • Stabilization • Object tracking • Image enhancement	• Windows • Linux • Solaris • Android	• GCC • Visual C++ • Visual Basic • Visual C# • Java and NetBeans • Sun Studio	30 days trial available • Moving target detection \$2,495 • Super resolution \$1,495 • Stabilization \$695 • Object tracking \$995	Depends on modules

Company	COTS Software	Web-page	Key Capabilities	Supported Platforms	Languages / Programming Environments	Notes	Price
Ilisys	IntelliViX-SDK	[R-19]	A collection of software modules that allow manufacturers to embed intelligence video analysis capability into their video surveillance products. It provides functions such as: <ul style="list-style-type: none">• Object detection and tracking• Object classification (human, vehicle or unknown)• Event detection	• Windows • Texas Instrument DaVinci chip	• C++	Outdated information as webpage was last updated in 2006	
intuVision	Video Analytics SDK	[R-20]	Provides real-time video object tracking and event detection functionalities such as <ul style="list-style-type: none">• Motion detection• Object classification (human, vehicle or animal)• People counting• Leaving objects, objects taken	• NVIDIA GPU	• C	Offer free trial	
Carnegie Mellon University	VIVID Tracker Testbed	[R-21]	A set of basic template tracking algorithms and a framework for testing them <ul style="list-style-type: none">• The user first defines a region around an object of interest, from which the template is extracted, the algorithm then tracks the template/object at each frame	• Windows	• C++ (DLL & source code) • Matlab (MEX)	Developed by students and academic researchers	Free
SRI International	Acadia II Embedded Video Processors and SDK	[R-22]	Support embedded surveillance and portable processing for weapon-mounted sights, handheld target finders, etc. Support applications including: <ul style="list-style-type: none">• AdapTrac tracks targets that move• Stabilization reduces video defects from uncontrolled camera motion• Motion detection & object tracking	ARM11 Quad MPCore • System-on-a-chip • System-on-a-module	• XML (for configuration of Acadia II) • Compact Vision System	Suitable for deployment on Small Arms	

3 REFERENCES

- R-1 VIRAT Video Datasets, <http://www.viratdata.org/> [Accessed 11 October 2013].
- R-2 PETS Video Datasets, <http://ftp.pets.rdg.ac.uk/> [Accessed 11 October 2013].
- R-3 Sentient Kestrel, <http://www.sentientvision.com/> [Accessed 11 October 2013].
- R-4 Deep Vision, <http://www.deepvision.ca/> [Accessed 11 October 2013].
- R-5 iOmniscient iQ series,
http://iomniscient.com/index.php?option=com_content&view=article&id=56&Itemid=53 [Accessed 11 October 2013].
- R-6 Cognika Pegasus, <http://www.cognika.com/index.php#content2> [Accessed 11 October 2013].
- R-7 Cognika Perseus, <http://www.cognika.com/index.php#post> [Accessed 11 October 2013].
- R-8 Fraunhofer IOSB ABUL,
<http://www.iosb.fraunhofer.de/servlet/is/28468/ABUL.pdf?command=downloadContent&filename=ABUL.pdf> [Accessed 11 October 2013].
- R-9 HP Surveillance, <http://www.autonomy.com/offerings/information-analytics/video-surveillance> [Accessed 11 October 2013].
- R-10 ObjectVideo, <http://www.objectvideo.com/> [Accessed 11 October 2013].
- R-11 Eptascape EptaAnalytics, <http://www.eptascape.com/products/eptaAnalytics.html> [Accessed 11 October 2013].
- R-12 Aimetis VE Series, <http://www.aimetis.com/Embedded/default.aspx> [Accessed 11 October 2013].

- R-13 Honeywell Active Alert Video Analytics,
<http://www.honeywellvideo.com/products/ias/va/160966.html> [Accessed 11 October 2013].
- R-14 IntelliVision Intelligent Video Analytics, <http://www.intelli-vision.com/products/intelligent-video-analytics> [Accessed 11 October 2013].
- R-15 MotionDSP Ikena ISR, <http://www.motiondsp.com/products/ikena-isr> [Accessed 11 October 2013].
- R-16 OpenCV, <http://www.opencv.org/> [Accessed 11 October 2013].
- R-17 Charles River Analytics VisionKit, <https://www.cra.com/commercial-solutions/computer-vision-components.asp> [Accessed 11 October 2013].
- R-18 2D3 Tungsten Media Toolkit, <http://www.2d3.com/product/?v=21> [Accessed 11 October 2013].
- R-19 Illisis IntelliVIX-SDK, <http://www.illisis.com/eng/panovix03.html> [Accessed 11 October 2013].
- R-20 intuVision Video Analytics SDK, <http://www.intuvisiontech.com/products/sdk.php> [Accessed 11 October 2013].
- R-21 VIVID Tracker Testbed, <http://vision.cse.psu.edu/data/vividEval/software.html> [Accessed 11 October 2013].
- R-22 SRI International Acadia II Embedded Video Processors,
<http://www.sri.com/engage/products-solutions/acadia-ii-embedded-video-processors> [Accessed 11 October 2013].